SAND2000-8566
Unlimited Release
Printed March 2000

Orthogonal Tensor Decompositions

T. G. Kolda*
Computational Sciences and Mathematics Research Department
Sandia National Laboratories
Livermore, CA 94551-9217

ABSTRACT

We explore the orthogonal decomposition of tensors (also known as multi-dimensional arrays or n-way arrays) using two different definitions of orthogonality. We present numerous examples to illustrate the difficulties in understanding such decompositions. We conclude with a counterexample to a tensor extension of the Eckart-Young SVD approximation theorem by Leibovici and Sabatier [*Linear Algebra Appl.* 269(1998):307–329].

Keywords: tensor decomposition, singular value decomposition, principal components analysis, multidimensional arrays.

*Email: tgkolda@ca.sandia.gov. This work was supported by the Applied Mathematical Sciences Research Program, Office of Energy Research, U.S. Department of Energy, under contracts DE–AC05–96OR22464 with Lockheed Martin Energy Research Corporation and DE-AC04-94AL85000 with Sandia Corporation.

This page intentionally left blank.

1 Introduction

The problem of decomposing tensors (also called n-way arrays or multidimensional arrays) is approached in a variety of ways by extending the singular value decomposition (SVD), principal components analysis (PCA), and other methods to higher orders; see, e.g., [1, 2, 6, 7, 8, 9, 10, 11, 12]. Tensor decompositions are most often used for multimode statistical analysis and clustering, but may also be used for compression of multidimensional arrays in ways similar to using a low-rank SVD for matrix compression. For example, color images are often stored as a sequence of RGB triplets, i.e., as separate red, green and blue overlays. An $m \times n$ pixel RGB image is represented by an $m \times n \times 3$ array, and a collection of p such images is an $m \times n \times 3 \times p$ array and can be compressed by a low-rank approximation.

The notation and basic properties of tensors are set forth in §2. Several definitions of orthogonality and several rank orthogonal decompositions for tensors are given in §3. Computational issues for orthogonal decompositions are discussed in §4. Finally in§5, we present a counterexample to Leibovici and Sabatier's extension of the well-known Eckart-Young SVD approximation theorem to tensors [10].

2 Tensors

Let A be an $m_1 \times m_2 \times \cdots \times m_n$ *tensor* over $\Re$. The *order* of A is n. The jth *dimension* of A is m_j. An element of A is specified as

$$A_{i_1 i_2 \cdots i_n},$$

where $i_j \in \{1, 2, \ldots, m_j\}$ for $j = 1, \ldots, n$. The set of all tensors of size $m_1 \times m_2 \times \cdots \times m_n$ is denoted by $\mathcal{T}(m_1, m_2, \ldots, m_n)$. The shorthand $\mathcal{T}_n$ may be used when only the order needs to be specified, or just $\mathcal{T}$ may be used when the order and dimensions are unambiguous.

Let $A, B \in \mathcal{T}(m_1, m_2, \ldots, m_n)$. The *inner product*[1] of A and B is defined as

$$A \cdot B \equiv \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} A_{i_1 i_2 \cdots i_n} B_{i_1 i_2 \cdots i_n}.$$

Correspondingly, the *norm* of A, $\|A\|$, is defined as

$$\|A\|^2 \equiv A \cdot A = \sum_{i_1=1}^{m_1} \sum_{i_2=1}^{m_2} \cdots \sum_{i_n=1}^{m_n} A^2_{i_1 i_2 \cdots i_n}.$$

We say A is a *unit* tensor if $\|A\| = 1$.

Example 1 Let $x, y \in \mathcal{T}(m)$; that is, x, y are vectors in $\Re^m$. Then $x \cdot y = x^T y$ where the superscript T denotes transpose. □

[1] In [10], the term is "contracted product" and the notation is $\langle A, B \rangle$.

Tensors of different orders may also be multiplied as follows. Suppose $C \in \mathcal{T}(m_1, \ldots, m_{j-1}, m_{j+1}, \ldots, m_n)$ is a tensor of order $n-1$ (note that m_j is missing). Then the *contracted product*[2] of A and C is a vector of length m_j, and its i_jth ($1 \leq i_j \leq m_j$) element is defined as

$$(A \cdot C)_{i_j} \equiv \sum_{i_1=1}^{m_1} \cdots \sum_{i_{j-1}=1}^{m_{j-1}} \sum_{i_{j+1}=1}^{m_{j+1}} \cdots \sum_{i_n=1}^{m_n} A_{i_1 \cdots i_{j-1} i_j i_{j+1} \cdots i_n} C_{i_1 \cdots i_{j-1} i_{j+1} \cdots i_n}.$$

Note that the same notation is used for both contracted and inner products—the difference is in the order of the tensors being multiplied.

Example 2 Suppose $A \in \mathcal{T}(m_1, m_2)$ is a tensor of order two, i.e., A is a matrix. If $b \in \mathcal{T}(m_1)$, then $A \cdot b = A^T b$ in matrix notation. Similarly, if $c \in \mathcal{T}(m_2)$, then $A \cdot c = Ac$. □

A *decomposed tensor* is a tensor $U \in \mathcal{T}(m_1, m_2, \ldots, m_n)$ that can be written as

$$U = u^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(n)}, \tag{1}$$

where $\otimes$ denotes the outer product and each $u^{(j)} \in \Re^{m_j}$ for $j = 1, \ldots, n$. The vectors $u^{(j)}$ are called the *components* of U. In this case,

$$U_{i_1 i_2 \cdots i_n} = u^{(1)}_{i_1} u^{(2)}_{i_2} \cdots u^{(n)}_{i_n}.$$

A decomposed tensor is a tensor of rank one for all the definitions of rank that we present in the next section. Decomposed tensors form the building blocks for tensor decompositions. The set of all decomposed tensors of size $m_1 \times m_2 \times \cdots \times m_n$ is denoted by $\mathcal{D}(m_1, m_2, \ldots, m_n)$ with shorthands analogous to $\mathcal{T}$.

Lemma 1 *Let $U, V \in \mathcal{D}$ where U is defined as in (1) and V is defined by*

$$V = v^{(1)} \otimes v^{(2)} \otimes \cdots \otimes v^{(n)}. \tag{2}$$

Then

$$\text{(a)} \quad U \cdot V = \prod_{j=1}^{n} u^{(j)} \cdot v^{(j)}, \qquad \text{(b)} \quad \|U\| = \prod_{j=1}^{n} \|u^{(j)}\|_2,$$

and, (c) $U + V \in \mathcal{D}$ if and only if all but at most one of the components of U and V are equal (within a scalar multiple).

Proof. Items (a) and (b) follow directly from the definitions. For item (c), consider $U, V \in \mathcal{D}$ such that $n-1$ components are equal, i.e., $w^{(i)} = v^{(i)}$ for all $i = 2, \ldots, n$. Then $W \equiv U + V$ can be written as

$$W = w^{(1)} \otimes u^{(2)} \otimes \cdots \otimes u^{(n)},$$

[2] In [10], the notation $A \,..\, C$ is used for contracted products.

where $w^{(1)} = u^{(1)} + v^{(1)}$, so the "if" statement of (c) is true. Next we show the "only if" statement of (c). First consider the special case where $n = 2$, $m_1 = m_2 = 2$,

$$U \equiv \begin{bmatrix} a \\ b \end{bmatrix} \otimes \begin{bmatrix} c \\ d \end{bmatrix}, V \equiv \begin{bmatrix} e \\ f \end{bmatrix} \otimes \begin{bmatrix} g \\ h \end{bmatrix},$$

and $W \equiv U + V \in \mathcal{D}$. Since $W \in \mathcal{D}$, we can write it as

$$W \equiv \begin{bmatrix} p \\ q \end{bmatrix} \otimes \begin{bmatrix} r \\ s \end{bmatrix}.$$

Then, we have

$$\begin{aligned} pr &= ac + eg, && (3) \\ ps &= ad + eh, && (4) \\ qr &= bc + fg, && (5) \\ qs &= bd + fh. && (6) \end{aligned}$$

Dividing (3) by (5) and (4) by (6) yields two ratios for p/q, and setting those equals gives

$$\frac{ac + eg}{bd + fh} = \frac{bc + fg}{ad + eh}. \tag{7}$$

Cross-multiplying and simplifying (7) finally yields

$$(af - be)(ch - dg) = 0.$$

In other words, either $u^{(1)} = v^{(1)}$ or $u^{(2)} = v^{(2)}$ (within a scalar multiple). So, all but at most one of the components of U and V must match if $W \in \mathcal{D}$. This argument can be extended to arbitrary n and m_j. □

Without loss of generality, we assume that the components of unit decomposed tensors are each unit vectors.

Although we have shown that for two decomposed tensors to be combined to one decomposed tensor they must match in all but at most one component, the same is not necessarily true when combining three or more decomposed tensors., as shown in the next example.

Example 3 Consider the following example. Let $a, b \in \Re^m$ with $a \perp b$ and $||a|| = ||b|| = 1$. Define $c = \frac{1}{\sqrt{2}}(a + b)$, and

$$U_1 = a \otimes a \otimes a, \quad U_2 = a \otimes b \otimes c, \quad U_3 = a \otimes c \otimes b.$$

Then the sum of theses three decomposed tensors can be rewritten as the sum of two despite the fact that they only match in one component:

$$U_1 + U_2 + U_3 = \sqrt{\frac{3}{2}} \left(V_1 + V_2 \right),$$

where

$$V_1 = a \otimes d \otimes b, \quad V_2 = a \otimes e \otimes b,$$

with

$$d = \sqrt{\frac{2}{3}}\, a + \sqrt{\frac{1}{3}}\, b, \quad e = \sqrt{\frac{2}{3}}\, c + \sqrt{\frac{1}{3}}\, b.$$

This is the result of splitting U_2 into two pieces based on the third component. $\square$

Lemma 2 *Let $U \in \mathcal{D}$ as defined in (1) and $A \in \mathcal{T}$. Then*

$$A \cdot U = (A \cdot u^{(1)} \otimes \cdots \otimes u^{(j-1)} \otimes u^{(j+1)} \otimes \cdots \otimes u^{(n)}) \cdot u^{(j)}.$$

Proof. Follows from the definitions. $\square$

3 Orthogonal Rank Decompositions

3.1 Notions of Orthogonality

Let $U, V \in \mathcal{D}$ be defined as in (1) and (2) respectively with $\|U\| = \|V\| = 1$. We say that U and V are *orthogonal* ($U \perp V$) if

$$U \cdot V = \prod_{j=1}^{n} u^{(j)} \cdot v^{(j)} = 0.$$

We say that U and V are *completely orthogonal* ($U \perp_{\mathrm{c}} V$) if for every $j = 1 \ldots, n$,

$$u^{(j)} \perp v^{(j)}.$$

We say that U and V are *strongly orthogonal* ($U \perp_{\mathrm{s}} V$) if $U \perp V$ and for every $j = 1, \ldots, n$,

$$u^{(j)} = \pm\, v^{(j)} \text{ or } u^{(j)} \perp v^{(j)}.$$

From the definition it follows that at least one pair must satisfy $u^{(j)} \perp v^{(j)}$.

Lemma 3 *Let the decomposed tensors U and V of order n be defined as in (1) and (2) respectively. Then*

$$U \perp_{\mathrm{c}} V \Rightarrow U \perp_{\mathrm{s}} V \Rightarrow U \perp V.$$

3.2 Rank Decompositions

Our goal is to express a tensor $A \in \mathcal{T}$ as a weighted sum of decomposed tensors,

$$A = \sum_{i=1}^{r} \sigma_i U_i, \tag{8}$$

where $\sigma_i > 0$ for $i = 1, \ldots, r$ and each $U_i \in \mathcal{D}$ and $\|U_i\| = 1$ for $i = 1, \ldots, r$.

- The *rank* of A, denoted $\text{rank}(A)$, is defined to be the minimal r such that A can be expressed as in (8). The decomposition is called the *rank decomposition.*

- The *orthogonal rank* of A, denoted $\text{rank}_{\perp}(A)$, is defined to be the minimal r such that A can be expressed as in (8) and $U_i \perp U_j$ for all $i \neq j$. The decomposition is called the *orthogonal rank decomposition.*

- The *strong orthogonal rank* of A, denoted $\text{rank}_{\perp_s}(A)$, is defined to be the minimal r such that A can be expressed as in (8) and $U_i \perp_s U_j$ for all $i \neq j$. The decomposition is called the *strong orthogonal rank decomposition.*[3]

As reported in [10], the definition of rank is due to Kruskal [8] and others, and the definitions of orthogonal and strong orthogonal rank is due to Franc [4]. The general *decomposition*, *orthogonal decomposition*, and *strong orthogonal decomposition* satisfy the orthogonality constraints (if any) but are not necessarily minimal in terms of r.

A slightly different notion of rank that depends on a type of strong orthogonal decomposition is the *combinatorial orthogonal rank*, denoted $\text{rank}_{\perp_c}(A)$. It is defined as the minimal r such that A can be written as

$$\sum_{i_1=1}^{r} \sum_{i_2=1}^{r} \cdots \sum_{i_n=1}^{r} \sigma_{i_1 i_2 \cdots i_n} \, u_{i_1}^{(1)} \otimes u_{i_2}^{(2)} \otimes \cdots u_{i_n}^{(n)}, \tag{9}$$

where

$$U_i = u_i^{(1)} \otimes u_i^{(2)} \otimes \cdots \otimes u_i^{(n)},$$

$U_i \perp_c U_j$ for all $i \neq j$, $1 \leq i, j \leq r$, and $\|U_i\| = 1$, $1 \leq i \leq r$. The decomposition (9) is the result of combining the components of the U_i's in every possible way and is called the *combinatorial orthogonal rank decomposition.* In this case, there are n^r scalar multiples (i.e., σ-values) that are involved rather than just r as in the other decompositions. This definition is a variation of the definition of rank for the higher-order SVD (HOSVD) by De Lathauwer [2] .

Lemma 4 *The rank, orthogonal rank, strong orthogonal rank, and combinatorial orthogonal rank decompositions are each equivalent to the SVD for tensors of order two.*

Proof. This follows from the properties of the SVD (c.f., [5]). $\square$

Example 4 Let $a, b \in \Re^m$ with $a \perp b$, and let $\sigma_1 > \sigma_2 > \sigma_3 > 0$. Define $A \in \mathcal{T}(m, m, m)$ as

$$A = \sigma_1 \underbrace{a \otimes b \otimes b}_{U_1} + \sigma_2 \underbrace{b \otimes b \otimes b}_{U_2} + \sigma_3 \underbrace{a \otimes a \otimes b}_{U_3}. \tag{10}$$

[3] In [10], the terms "free orthogonal rank" and "free rank decomposition" are used rather than "strong orthogonal rank" and "strong orthogonal rank decomposition".

Note that $U_i \perp_s U_j$ for all $i \neq j$, so (10) is a strong orthogonal decomposition of A. Furthermore, A cannot be expressed as the sum of fewer weighted strong orthogonal decomposed tensors, so the strong orthogonal rank of A is three. Observe that A can also be expressed as

$$A = \hat{\sigma}_1 \underbrace{\hat{b} \otimes b \otimes b}_{\hat{U}_1} + \hat{\sigma}_2 \underbrace{\hat{b} \otimes a \otimes b}_{\hat{U}_2} + \hat{\sigma}_3 \underbrace{\hat{a} \otimes a \otimes b}_{\hat{U}_3}, \tag{11}$$

where

$$\hat{\sigma}_1 = \sqrt{\sigma_1^2 + \sigma_2^2} \ , \quad \hat{\sigma}_2 = \frac{\sigma_1 \ \sigma_3}{\hat{\sigma}_1} \ , \quad \hat{\sigma}_3 = \frac{\sigma_2 \ \sigma_3}{\hat{\sigma}_1} \ ,$$

$$\hat{a} = \frac{\sigma_1 \ a + \sigma_2 \ b}{\hat{\sigma}_1} \ , \quad \text{and} \quad \hat{b} = \frac{\sigma_2 \ a - \sigma_1 \ b}{\hat{\sigma}_1} \ .$$

Since $\hat{a} \perp \hat{b}$, we have $\hat{U}_i \perp_s \hat{U}_j$ for all $i \neq j$. Therefore (11) is also a strong orthogonal rank decomposition of A, and so the strong orthogonal rank decomposition is not unique. It follows immediately that the closely related combinatorial orthogonal rank decomposition is not unique. □

Example 5 Consider the tensor A as defined by (10); A can also be written as

$$A = \bar{\sigma}\bar{U} + \sigma_3 U_3, \tag{12}$$

where

$$\bar{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2} \quad \text{and} \quad \bar{U} = \frac{\sigma_1 \ a + \sigma_2 \ b}{\bar{\sigma}} \otimes b \otimes b.$$

Observe that $\bar{U} \perp U_3$; in fact, (12) is an orthogonal rank decomposition of A, and therefore the orthogonal rank of A is two. Alternatively from (11), we can express A as

$$A = \breve{\sigma}\breve{U} + \hat{\sigma}_3 \hat{U}_3, \tag{13}$$

where

$$\breve{\sigma} = \sqrt{\sigma_1^2 + \sigma_2^2 + \frac{\sigma_1^2 \sigma_3^2}{\sigma_1^2 + \sigma_2^2}} \quad \text{and} \quad \breve{U} = \hat{b} \otimes \frac{\hat{\sigma}_1 b + \hat{\sigma}_2 a}{\breve{\sigma}} \otimes b.$$

Note that $\breve{U} \perp hat U_3$, so () is also an orthogonal rank decomposition of A, so the orthogonal rank decomposition is not unique. (Two other orthogonal rank decompositions of A are possible as well.) □

Lemma 5 *Neither the orthogonal rank, strong orthogonal rank, nor combinatorial orthogonal rank decomposition is unique.*

Proof. See Examples 4 and 5. □

The lack of uniqueness in the various rank tensor decompositions is an important difference between tensor and matrix decompositions.

Example 6 We show how to 'orthogonalize' a tensor in a relatively simple situation. Suppose that we have an order three tensor $A \in \mathcal{T}(m_1, m_2, m_3)$ defined as follows:

$$A = \sigma_1 U + \sigma_2 V,$$

where $\sigma_1 \geq \sigma_2$ and,

$$\begin{aligned} U &= u^{(1)} \otimes u^{(2)} \otimes u^{(3)}, \\ V &= v^{(1)} \otimes v^{(2)} \otimes v^{(3)}, \end{aligned}$$

with $u^{(i)}, v^{(i)}$ unequal, non-orthogonal unit vectors in $\Re^{m_i}$ for $i = 1, 2, 3$.

For $i = 1, 2, 3$, we can decompose $v^{(i)}$ as

$$v^{(i)} = \alpha^{(i)} u^{(i)} + \hat{\alpha}^{(i)} \hat{u}^{(i)},$$

where

$$\begin{aligned} \alpha^{(i)} &= v^{(i)} \cdot u^{(i)}, \\ \hat{\alpha}^{(i)} &= \|v^{(i)} - \alpha^{(i)} u^{(i)}\|, \text{ and} \\ \hat{u}^{(i)} &= (v^{(i)} - \alpha^{(i)} u^{(i)}) / \hat{\alpha}^{(i)}. \end{aligned}$$

Then, we can rewrite A as

$$\begin{array}{rccc} A & = & (\sigma_1 + \sigma_2\ \alpha^{(1)}\alpha^{(2)}\alpha^{(3)}) & u^{(1)} \otimes u^{(2)} \otimes u^{(3)} \\ & + & \sigma_2\ \alpha^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)} & u^{(1)} \otimes u^{(2)} \otimes \hat{u}^{(3)} \\ & + & \sigma_2\ \alpha^{(1)}\hat{\alpha}^{(2)}\alpha^{(3)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes u^{(3)} \\ & + & \sigma_2\ \alpha^{(1)}\hat{\alpha}^{(2)}\hat{\alpha}^{(3)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes \hat{u}^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\alpha^{(2)}\alpha^{(3)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes u^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes \hat{u}^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\hat{\alpha}^{(2)}\alpha^{(3)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes u^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\hat{\alpha}^{(2)}\hat{\alpha}^{(3)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes \hat{u}^{(3)}. \end{array} \tag{14}$$

Equation (14) shows that $\text{rank}_{\perp_s}(A) \leq 8$. Because of the way U and V were chosen (components neither equal nor orthogonal), equation (14) is a strong orthogonal rank decomposition of A, and $\text{rank}_{\perp_s}(A) = 8$. (From Equation (14), we can also deduce that $\text{rank}_{\perp_c}(A) = 2$.) This is not, however, an orthogonal rank decomposition. Combining each pair of lines in (14), we get

$$\begin{array}{rccl} A & = & \sqrt{\gamma^2 + \hat{\gamma}^2} & u^{(1)} \otimes u^{(2)} \otimes (\gamma u^{(3)} + \hat{\gamma}\hat{u}^{(3)}) / \sqrt{\gamma^2 + \hat{\gamma}^2} \\ & + & \sigma_2\ \alpha^{(1)}\hat{\alpha}^{(2)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\alpha^{(2)} & \hat{u}^{(1)} \otimes u^{(2)} \otimes v^{(3)} \\ & + & \sigma_2\ \hat{\alpha}^{(1)}\hat{\alpha}^{(2)} & \hat{u}^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)}. \end{array} \tag{15}$$

where

$$\gamma = \sigma_1 + \sigma_2\ \alpha^{(1)}\alpha^{(2)}\alpha^{(3)} \text{ and } \hat{\gamma} = \sigma_2\ \alpha^{(1)}\alpha^{(2)}\hat{\alpha}^{(3)}.$$

Finally, combining the last two lines of (15), we arrive at an orthogonal rank decomposition,

$$\begin{array}{rccl} A & = & \sqrt{\gamma^2+\hat{\gamma}^2} & u^{(1)} \otimes u^{(2)} \otimes (\gamma u^{(3)} + \hat{\gamma}\hat{u}^{(3)})/\sqrt{\gamma^2+\hat{\gamma}^2} \\ & + & \sigma_2\, \alpha^{(1)}\hat{\alpha}^{(2)} & u^{(1)} \otimes \hat{u}^{(2)} \otimes v^{(3)} \\ & + & \sigma_2\, \hat{\alpha}^{(1)} & \hat{u}^{(1)} \otimes v^{(2)} \otimes v^{(3)}, \end{array}$$

so $\mathrm{rank}_\perp(A) = 3$. Note that combining vectors from (14) in different order would have resulted in a different orthogonal rank decomposition. □

Theorem 1 ([10]) *For a given tensor A,*

$$\mathrm{rank}(A) \leq \mathrm{rank}_\perp(A) \leq \mathrm{rank}_{\perp_s}(A). \tag{16}$$

Proof. This follows from Lemma 3. □

Corollary 1 ([10]) *For any $A \in \mathcal{T}_2$,*

$$\mathrm{rank}(A) = \mathrm{rank}_\perp(A) = \mathrm{rank}_{\perp_s}(A) = \mathrm{rank}_{\perp_c}(A).$$

Proof. This follows from Lemma 4. □

Corollary 2 *For any order $n > 2$, there exists $A \in \mathcal{T}_n$ such that strict inequality holds in (16).*

Proof. An example of strict inequality for a tensor of order three ($n = 3$) is given in Example 6, and that example can be generalized to any order. □

In our discussion of rank decomposition, we did not present a *completely orthogonal decomposition*. In fact, we are not in general guaranteed that such a decomposition can be found. A completely orthogonal decomposition corresponds to a combinatorial orthogonal decomposition in which only the diagonal elements ($\sigma_{ii\cdots i}$) are nonzero; and so, in general, tensors cannot be *diagonalized*.

Corollary 3 ([10]) *If a tensor can be decomposed as the weighted sum of completely orthogonal decomposed tensors, then equality holds in (16).*

Proof. Follows from the definitions. □

Matrices (i.e., tensors of order two) are special cases that always have a completely orthogonal decomposition.

Corollary 4 *For any order $n > 2$, there exists $A \in \mathcal{T}_n$ such that A cannot be decomposed as the weighted sum of completely orthogonal tensors.*

Proof. See the construction of the decompositions of A in Example 6. $\square$

We now have several examples illustrating that the strong orthogonal rank and orthogonal rank decompositions are not unique. A partial 'fix' for lack of uniqueness is the following. Without loss of generality, assume that the σ_i's in (8) are always ordered so that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then define the *unique* (strong) orthogonal rank decomposition to be the (strong) orthogonal rank decomposition that has the largest possible σ_1, and given that choice for σ_1, has the largest possible σ_2, and so forth. This decomposition is unique in the sense that the weights are unique. The unit decomposed tensors are unique if and only if no two σ_i's are equal. A *unique* combinatorial orthogonal rank decomposition can be defined is a more complicated way by choosing the kth U so that

$$\sum_{i_1=1}^{k} \sum_{i_2=1}^{k} \cdots \sum_{i_n=1}^{k} \sigma_{i_1 i_2 \cdots i_n}^2,$$

is maximized.

Example 7 In Example 4, the unique strong orthogonal rank decomposition is given by (11). Similarly, in Example 5, the unique orthogonal rank decomposition is given by (5). $\square$

4 Greedy Tensor Decompositions

We present a method for generating a *greedy orthogonal decomposition.* Our goal is to compute a series of weighted decomposed tensors such that

$$A = \sum_{i=1}^{p} \sigma_i U_i,$$

where $U_i \perp U_j$ for all $i \neq j$ and $\|U_i\| = 1$ for all i. We do not yet make any claims as to whether or not this greedy orthogonal decomposition yields a orthogonal rank decomposition.

In the greedy orthogonal decomposition, the $\{\sigma, U\}$ pairs are computed iteratively as follows. Define the kth residual tensor

$$R_k \equiv A - \sum_{i=1}^{k} \sigma_i U_i,$$

with $R_0 = A$, and let the set of tensors $\mathcal{U}_k$ be defined as

$$\mathcal{U}_k = \{U_1, U_2, \ldots, U_k\},$$

with $\mathcal{U}_0 = \emptyset$. Our goal is to find the best rank-1 approximation to the current residual subject to orthogonality constraints; that is, we wish to solve

$$\min \ f_k(\sigma, U) \equiv \|R_k - \sigma U\|^2, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp \mathcal{U}_k.$$

We can rewrite f_k as

$$f_k(\sigma, U) = \|R_k\|^2 - 2\sigma R_k \cdot U + \sigma^2 \|U\|^2.$$

At the solution, we have

$$\frac{\partial f_k}{\partial \sigma} \equiv -2R_k \cdot U + 2\sigma \|U\|^2 = 0,$$

so we can solve for σ and conclude that minimizing f_k is the same as solving

$$\max \ R_k \cdot U, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp \mathcal{U}_k. \tag{17}$$

Define U_{k+1} to be the solution of (17), and let $\sigma_{k+1} = R_k \cdot U_{k+1}$.

A *greedy strong orthogonal decomposition* can be similarly described, and reduces to solving

$$\max \ R_k \cdot U, \ \text{ s.t. } \ U \in \mathcal{D}, \ \ \|U\| = 1, \ \ U \perp_{\mathrm{s}} \mathcal{U}_k. \tag{18}$$

Lemma 6 *The greedy (strong) orthogonal decomposition is finite.*

Proof. This is a consequence of the fact that there are at most $M = \prod_{j=1}^{n} m_j$ (strong) orthogonal decomposed tensors. $\square$

Solving (17) or (18) is a very challenging task. For example, in order to solve (17), we might use an *alternating least squares* (ALS) approach as follows. For $l = 1, \ldots, n$, fix all components of U but the lth, and solve

$$\max \ s \cdot u^{(l)}, \ \text{ s.t. } \ \|U\| = 1, \ \ U \perp \mathcal{U}_k$$

where

$$s = R_k \cdot u^{(1)} \otimes \cdots \otimes u^{(l-1)} \otimes u^{(l+1)} \otimes \cdots \otimes u^{(n)}.$$

The difficulty with this approach is in enforcing the constraints.

We may also construct a sort of greedy approach for the combinatorial orthogonal decomposition, but the subproblems are even more complicated in this case.

Zhang and Golub [12] explore various computational techniques when the tensor has a completely orthogonal decomposition, in which case the problem is much simpler. In [10], the RPVSCC method uses ALS to the the *modes*, i.e., the completely orthogonal decomposed tensors, and then fills in the values associated with the combinations of the components of the modes. De Lathauwer [2] presents several ALS methods for computing the HOSVD (a special type of strong orthogonal decomposition). Kroonenberg and Jan de Leeuw [7] propose an alternating least squares solution to (9) so that at each step an entire set $\{u_i^{(j)}\}_{i=1}^{m_j}$ is solved for some j while everything else is fixed. In other words, the method concentrates on one subspace at a time.

5 Approximation of a Tensor

The well-known Eckart-Young approximation theorem [3, 5] says that if the SVD of a matrix is given by

$$A = \sum_{i=1}^{r} \sigma_i u_i v_i^T,$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$, then the best rank-$k$ approximation is given by

$$A_k \equiv \sum_{i=1}^{r} \sigma_i u_i v_i^T.$$

A consequence of this result is that the SVD can be approximated via a greedy method which calculates each triplet $\{\sigma_i, u_i, v_i\}$ in sequence. Now we can ask whether or not the Eckart-Young theorem can be extended to tensor rank decompositions; i.e., is the best rank-k approximation of a tensor given by the sum of the first k terms in its rank decomposition? The relates directly to whether or not the greedy orthogonal, strong orthogonal, or combinatorial decompositions produce a corresponding rank decomposition.

In the case of the strong orthogonal rank decomposition, the answer is definitely no, contrary to the result stated in [10], as the following counterexample shows.

Example 8 Consider the strong orthogonal rank decomposition of a matrix $A \in \mathcal{T}(m, m, m)$ defined by

$$A = \sum_{i=1}^{6} \sigma_i U_i,$$

where the $\{\sigma_i, U_i\}$ pairs are defined as follows. Let the vectors $a, b, c, d \in \Re^m$ be two-by-two orthogonal, then let

$$\begin{aligned}
\sigma_1 &= 1.00, & U_1 &= a \otimes a \otimes a, \\
\sigma_2 &= 0.75, & U_2 &= b \otimes b \otimes b, \\
\sigma_3 &= 0.70, & U_3 &= a \otimes c \otimes d, \\
\sigma_4 &= 0.70, & U_4 &= a \otimes d \otimes c, \\
\sigma_5 &= 0.65, & U_5 &= b \otimes c \otimes d, \\
\sigma_6 &= 0.65, & U_6 &= b \otimes d \otimes c.
\end{aligned}$$

Note that $\sigma_3 U_3$ and $\sigma_5 U_5$ can be combined to form the decomposed tensor

$$\gamma_1 V_1 \equiv \sqrt{\sigma_3^2 + \sigma_5^2}\, \frac{\sigma_3 a + \sigma_5 b}{\sqrt{\sigma_3^2 + \sigma_5^2}} \otimes c \otimes d. \tag{19}$$

Similarly, $\sigma_4 U_4$ and $\sigma_5 U_5$ can be combined to form

$$\gamma_2 V_2 \equiv \sqrt{\sigma_4^2 + \sigma_6^2}\, \frac{\sigma_4 a + \sigma_6 b}{\sqrt{\sigma_4^2 + \sigma_6^2}} \otimes d \otimes c. \tag{20}$$

But,

$$\gamma_1 = \gamma_2 \approx 0.9552 < \sigma_1 = 1,$$

so neither (19) nor (20) is the best rank one approximation to A; $A_1 \equiv \sigma_1 u_1$ is. However, the best strong orthogonal rank two approximation is given by

$$A_2 \equiv \gamma_1 V_1 + \gamma_2 V_2,$$

because $V_1 \perp_s V_2$ and

$$\gamma_1^2 + \gamma_2^2 = 1.825 > \sigma_1^2 + \sigma_2^2 = 1.5625.$$

Thus, we have a counterexample to any Eckart-Young type theorem for strong orthogonal rank decompositions. □

Example 8 can be reworked to show that the combinatorial orthogonal rank decomposition does not yield a best rank-k approximation either.

Example 9 Consider the tensor defined in Example 8. Let e and f be any vectors that are orthogonal to each other and also to a and b. We can express a combinatorial orthogonal rank decomposition of A as follows.

$$A = \sum_{i_1=1}^{4} \sum_{i_2=1}^{4} \sum_{i_3=1}^{4} \bar{\sigma}_{i_1 i_2 i_3} \bar{u}_{i_1}^{(1)} \otimes \bar{u}_{i_2}^{(2)} \otimes \bar{u}_{i_3}^{(3)},$$

where

$$\begin{array}{rclcrcl} \bar{U}_1 & = & a \otimes a \otimes a, & & \bar{U}_3 & = & e \otimes c \otimes d, \\ \bar{U}_2 & = & b \otimes b \otimes b, & & \bar{U}_4 & = & f \otimes d \otimes c, \end{array}$$

and the only non-zero $\bar{\sigma}$'s are

$$\bar{\sigma}_{111} = \sigma_1, \ \ \bar{\sigma}_{222} = \sigma_2, \ \ \bar{\sigma}_{133} = \sigma_3, \ \ \bar{\sigma}_{233} = \sigma_4, \ \ \bar{\sigma}_{144} = \sigma_5, \ \ \bar{\sigma}_{244} = \sigma_6.$$

So, $\mathrm{rank}_{\perp_c}(A) = 4$. The best combinatorial orthogonal rank-1 approximation to A is $\bar{A}_1 = \bar{\sigma}_{111}\bar{U}_1 = \sigma_1 U_1$ (the same as the best strong orthogonal rank-1 approximation). But, the best combinatorial orthogonal rank-2 approximation is yielded by

$$\bar{A}_2 = \sum_{i_1=1}^{2} \sum_{i_2=1}^{2} \sum_{i_3=1}^{2} \bar{\gamma}_{i_1 i_2 i_3} \bar{v}_{i_1}^{(1)} \otimes \bar{v}_{i_2}^{(2)} \otimes \bar{v}_{i_3}^{(3)}.$$

Here

$$\bar{V}_1 \equiv V_1 \quad \text{and} \quad \bar{V}_2 \equiv g \otimes d \otimes c,$$

where g is some vector orthogonal to $v_1^{(1)}$, and the only nonzero $\bar{\gamma}$'s are $\bar{\gamma}_{111} = \gamma_1$ and $\bar{\gamma}_{122} = \gamma_2$. □

Example 3 shows that it is possible to add an orthogonal decomposed tensor to a sum without increasing its rank ($U_1 + U_2$ has rank 2 as does $U_1 + U_2 + U_3$). This is contrary to a fundamental assumption used in the proof of Theorem 2 in [10]. So whether or not the Eckart-Young SVD approximation theorem can be extended to the orthogonal rank decomposition is still an open question.

Conjecture 1 (Eckart-Young extended) *Let the unique orthogonal rank decomposition of a tensor A be given as in (8) and assume that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r$. Then the best orthogonal rank p $(p < r)$ approximation to A satisfies*

$$\min_{\mathrm{rank}_\perp A_p = p} \|A - A_p\|^2 = \sum_{i=p+1}^{r} \sigma_i^2$$

and is given by

$$A_p \equiv \sum_{i=1}^{p} \sigma_i u_i.$$

6 Conclusions

There are multiple ways to orthogonally decompose tensors, depending both on the definition of orthogonality as well as on the definitions of decomposition and rank. An Eckart-Young type of best rank-k approximation theorem for tensors continues to elude our investigations but can perhaps eventually be attained by using a different norm or yet other definitions of orthogonality and rank.

Computing orthogonal tensor decomposition is a challenge as well. Most methods are variations on ALS, a method which can be very slow to converge, although recently several authors (c.f., [2, 12]) have presented new ideas.

Acknowledgements

The author wishes to thank Dianne O'Leary and Paul Boggs for their helpful advice in the preparation of this manuscript.

References

[1] J. D. Carroll and J.-J. Chang, *Analysis of individual differences in multidimensional scaling via an n-way generalization of "Eckart-Young" decomposition*, Psychometrika, 35 (1970), pp. 283–319.

[2] L. de Lathauwer, *Signal Processing Based on Multilinear Algebra*, PhD thesis, Katholieke Universiteit Leuven, Belgium, 1997.

[3] C. Eckart and G. Young, *The approximation of one matrix by another of lower rank*, Psychometrika, 1 (1936), pp. 211–218.

[4] A. Franc, *Etude Algébrique des Multitableaux: Apports de l'Algébre Tensorielle*, Thèse de Doctorat, Spécialité Statistiques, Univ. de Montpellier II, Montpellier, 1992.

[5] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The Johns Hopkins University Press, Baltimore, 3rd ed., 1996.

[6] A. Kapteyn, H. Neudecker, and T. Wansbeek, *An approach to n-mode components analysis*, Psychometrika, 51 (1986), pp. 269–275.

[7] P. M. Kroonenberg and J. de Leeuw, *Principal component analysis of three-mode data by means of alternating least squares algorithms*, Psychometrika, 45 (1980), pp. 69–97.

[8] J. B. Kruskal, *Three-way arrays: Rank and uniqueness of trilinear decompositions, with application to arithmetic complexity and statistics*, Linear Algebra Appl., 18 (1977), pp. 95–138.

[9] J. B. Kruskal, *Rank, decomposition, and uniqueness for 3-way and n-way arrays*, in Multiway Data Analysis, R. Coppi and S. Bolasco, eds., Elsevier Science Publishers B.V., North-Holland, 1989, pp. 7–18.

[10] D. Leibovici and R. Sabatier, *A singular value decomposition of a k-way array for principal component analysis of multiway data, PTA-k*, Linear Algebra Appl., 269 (1998), pp. 307–329.

[11] L. R. Tucker, *Some mathematical notes on three-mode factor analysis*, Psychometrika, 31 (1966), pp. 279–311.

[12] T. Zhang and G. H. Golub, *Rank-one approximation to high order tensors.* In review, 2000.